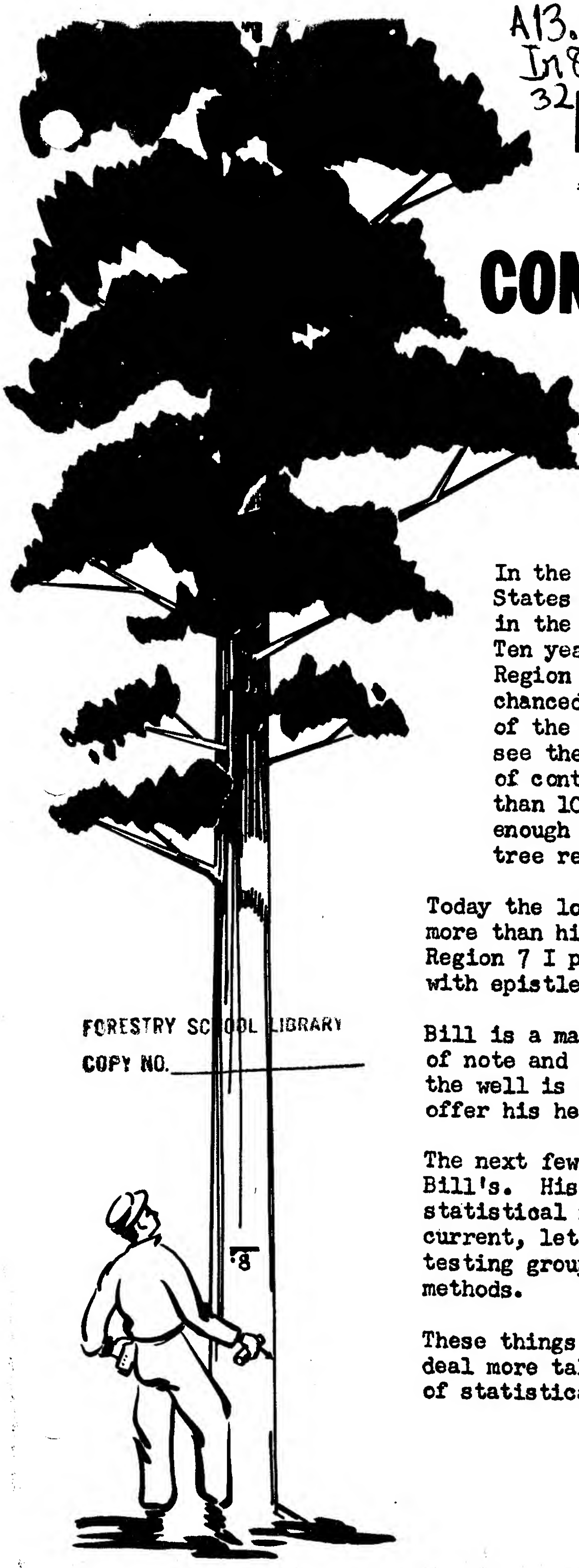

FORESTS CONTROL by CONTINUOUS INVENTORY

"---you shall not muzzle an ox when it is
treading out the grain."
First Corinthians 9:9

Milwaukee, Wis. November, 1956 No. 32

MORE TESTING - LESS TALKING

In the old Philadelphia Region 7 of the United States Forest Service, Bill Barton lives and works in the field of state and private cooperation. Ten years ago he worked with me in Milwaukee, Region 9. His well-deserved promotional transfer chanced to come at the time the industrial phase of the work began to increase. No one could foresee then that industry would buy millions of acres of continuous forest inventory controls in less than 10 years. Not one of us had imagination enough to foresee the mark sensing of individual tree records on almost a million IBM cards.

Today the load is heavy, and Barton, too, is carrying more than his share, for besides his home work in Region 7 I pester him with problems and pepper him with epistles on both statistics and C.F.I. with I.B.M.

Bill is a mathematician and a practicing statistician of note and I am not. I "have nothing to draw with and the well is deep," but he has been kind enough to offer his help freely.

The next few issues of the newsletter will largely be Bill's. His letters begin with this one on simplified statistical methods. Subsequent, although not concurrent, letters will cover more advanced phases of testing grouped and ungrouped data by hand and machine methods.

These things are important for there has been a great deal more tall talking than true testing in this field of statistical check methods.

CAL STOTT

The Forester

FOREST SERVICE U.S. DEPARTMENT OF AGRICULTURE

PRINCIPLES OF STATISTICAL ANALYSIS

Statistics is not really difficult. The amount of it that is useful to a forester is quite simple and quite logical. It is used to save time and effort in estimating amounts.

An absolutely accurate amount would be obtained by counting each individual, such as each tree on a specific tract. Since we do not usually need to know the amount exactly to the tree, board foot, or other unit, but rather we wish to have an estimated amount within some plus or minus percent of error from the real total, we use statistics to save a great deal of effort and time. With statistics we can control a survey--that is, use enough, but not more than enough, effort--to get an estimate to the degree of accuracy we wish to have.

If our population (meaning the stuff we are to find the amount of) is uniform, samples will be uniform and only a few will be needed to give a good estimate. We need some way to express uniformity, or the lack of it. There are many ways in which this might be done, but following a period of trial and error one method has evolved as best for most purposes. Standardization of the way in which to express the degree of uniformity laid the foundation of the science of statistics. Although there are experts (mathematicians who still specialize in the development of statistical methods) who are still trying out one thing and another, most of us are concerned with applying standardized statistical methods to practical uses and you may as well be one of us.

Mean

First come the definitions of simple things with complicated names. One of these is the MEAN, arithmetic average, or sometimes it is called just plain average. The statistical term is MEAN. It goes by the symbol (M) and is found by adding the amounts of all records and dividing this sum by the number of records, in the following manner.

	Individual Record (r)
	6
	7
	3
	2
Total	18

divided by 4 records equals 4.5 -- the average or MEAN (M).

Deviation

The next statistical term we shall wish to know is the DEVIATION. This is represented by the symbol (x). It is the difference between an individual record (r) and the MEAN (M). To carry our example to this step, we have the figures below.

Record (r)	Mean (M)	Deviation (x)
6	4.5	1.5
7	4.5	2.5
3	4.5	-1.5
2	4.5	-2.5

Standard Deviation

We now have the basic information inherent in the records. To show some measure of the uniformity of our population, we could just add up the deviations (disregarding plus or minus sign), find the average deviation and compare the result with the MEAN. This has been tried, but is not considered as good a figure to use as the root-mean-square of the deviations. This is found by (1) squaring each deviation, (2) finding the total of the squared deviations, (3) dividing by the number of records to get the average (mean) of the squares of the deviations and (4) extracting the square root of this figure. In other words it is the square root of the mean of the squares of the deviations. This is found as shown below:

Record (r)	Mean (M)	Deviation (x)	$(\text{Deviation})^2$ $(x)^2$
6	4.5	1.5	2.25
7	4.5	2.5	6.25
3	4.5	-1.5	2.25
2	4.5	-2.5	6.25
			4) 17.00
			4.25

Square root of 4.25 equals 2.061

This figure is known as the STANDARD DEVIATION, with the symbol σ. It is the basic calculated figure of statistics.

Coefficient of Variation

The standard deviation (σ) is still just a figure. Until it is compared with something it has no meaning. So let's divide the standard deviation (σ) by the mean (M), the logical figure with which to compare it.

$$\frac{\sigma}{M} = \frac{2.061}{4.5} \text{ equals } 0.458$$

When we express this figure as a percent, instead of a decimal, we call it the COEFFICIENT OF VARIATION represented by the symbol (c).

c equals 45.8%

Summary of Terms

On the basis of the few samples, we have developed certain figures which describe the character of our population, timber stand or stuff that we are seeking to find the amount of. These are:

MEAN (M) equals 4.5

STANDARD DEVIATION (σ) 2.061

COEFFICIENT OF VARIATION (c) 45.8%

Review

Our next step is to do something useful with this information, but suppose first we take a quick review with a new and slightly more complicated example.

Record (r)	Mean (M)	Deviation (x)	$(\text{Deviation})^2$ $(x)^2$
24	29	-5	25
31	29	2	4
25	29	-4	16
28	29	-1	1
17	29	-12	144
30	29	1	1
29	29	0	0
51	29	22	484
27	29	-2	4
28	29	-1	1
10) 290 = 29 (M)		0	10) 680 = 68 $(\sigma)^2$

Standard Deviation (σ) equals $\sqrt{68}$ or $\pm$ 8.246

$$\frac{(\sigma)}{M} \qquad \frac{8.246}{29} \text{ equals } 0.284 \text{ or the}$$

Coefficient of Variation (c) is 28.4%

****

Note that the sum of the deviations (x) is 0 in the case where we have used the exact MEAN. Now suppose we try a slight variation by using an assumed average (A) instead of the true mean (M). In our review example let's use 31 instead of 29, and see if we can eventually develop the exact same figures we had before. If we can work with an assumed average (A) and then simply adjust everything to be as it would be if we had used the exact mean (M), we can sometimes work with simple figures instead of rather complicated ones. The deviation to an assumed average (A) is designated ($\bar{x}$) to avoid confusion with the deviation (x) with respect to the true Mean (M).

Record	Assumed Average	Deviation from Assumed Average	(Deviation from Assumed Average)^2
(r)	(A)	($\bar{x}$)	$(\bar{x})^2$
24	31	-7	49
31	31	0	0
25	31	-6	36
28	31	-3	9
17	31	-14	196
30	31	-1	1
29	31	-2	4
51	31	20	400
27	31	-4	16
28	31	-3	9
10) 290		10) -20	10) 720
29 (M)		-2 (d)	72 $(s)^2$

We use the special symbol (d) to represent the algebraic sum of the deviations with respect to an assumed average (A) divided by the number of records. Likewise, we use the symbol $(s)^2$ to represent the square of the standard deviation with respect to the assumed average (A).

Note now that our deviations with respect to the assumed average (A) total −20, and that if we divide this number by the number of records (10), we get −2, which when subtracted from the <u>assumed average of 31</u> gives us our <u>true mean of 29.</u>

Assumed average	(A)	31
Sum ($\bar{x}$) / 10	(d)	−2
True Mean	(M)	29

You will also note, that with respect to our assumed average of 31, the standard deviation squared came out to 72 instead of 68 as when using the true mean. This may now be corrected to the true standard deviation squared by subtracting the square of (d) from $(s)^2$. Thus:

	$(s)^2$	72	
minus	$(d)^2$	4	
	$(\sigma)^2$	68	the true standard deviation squared.

Had the temporarily assumed average (A) been lower than the true mean (M), the exact same procedure would apply.

For example:

R Record	Assumed Average	Deviation	$(\text{Deviation})^2$
(r)	(A)	($\bar{x}$)	$(\bar{x})^2$
24	25	−1	1
31	25	6	36
25	25	0	0
28	25	3	9
17	25	−8	64
30	25	5	25
29	25	4	16
51	25	26	676
27	25	2	4
28	25	3	9
10) 290		10) + 40	10) 840
29 (M)	25 (A)	+ 4 (d)	84 $(s)^2$
	plus 4 (d) 1/		Minus 16 $(d)^2$ 1/
	29 (M)		68 $(\sigma)^2$

1/ Note we add or subtract (d) according to sign but we always <u>subtract</u> $(d)^2$.

Generally our data is much more complicated than that used in these examples. To treat it in the exact same way would be very laborious, so instead of working with individual records, we group them. When we collect them into about 20 evenly spaced groups, and work with group numbers from 1 to 20 we can do most of the multiplication and squaring in our heads. Our results are practically identical with the figures we would obtain by using each record individually. The saving in time is tremendous.

To make it easy to handle this data by groups, a STATISTICAL CHECK SHEET has been designed. A description of this sheet and its use in controlling sampling surveys is described in a separate booklet.

William Warren Barton
USDA-Forest Service
Eastern Region
Upper Darby, Pa.
Sept. 1956